

Increased complexity of wild-type adeno-associated virus–chromosomal junctions as determined by analysis of unselected cellular genomes

Horace R. Drew, Linda J. Lockett and Gerald W. Both

CSIRO Molecular and Health Technologies, North Ryde, NSW 2113, Australia

Correspondence
Gerald W. Both
gerry.both@csiro.au

Adeno-associated virus (AAV) undergoes preferential Rep-mediated integration into the AAVS1 region of human chromosome 19 during latent infection, at least in highly-selected cell cultures. However, integration at the level of the whole eukaryotic genome in unselected cells has not yet been monitored for AAV as it has been for retro- and lentiviruses. Here we have used ligation-mediated PCR (LMPCR) to monitor the formation of AAV–chromosome junctions within unselected genomic DNA after infection. Our analyses show that, in the absence of selection, the complexity of junction formation is much greater than for selected cells. Sequencing of more than 50 authentic LMPCR clones showed that AAV formed junctions with many different chromosomal sites via DNA micro-homologies that frequently involved GGTC motifs located within the AAV p5 element. One site at position 280 was preferred. Even greater complexity was found when unselected junctions identified by LMPCR were analysed by direct PCR amplification and cloning of genomic DNA. No clones containing AAV–AAVS1 chromosome 19 junctions were identified among the LMPCR clones, although they were readily obtained using chromosomal PCR primers, suggesting that junctions with AAVS1 constituted only a small portion of the total. Thus, we have identified an additional means by which AAV sequences may join to human chromosomes, although the detailed molecular mechanisms remain to be elucidated. These data may have implications for the design of new-generation AAV vectors.

Received 24 January 2007
Accepted 14 February 2007

INTRODUCTION

Adeno-associated virus type 2 (AAV) is a single-stranded parvovirus of humans that has a biphasic life cycle. In the presence of helpers such as adenovirus, AAV undergoes productive infection, but without help, infection remains latent and genome integration is catalysed by the viral Rep protein at a preferred site (AAVS1) on human chromosome 19q13.3-qter (Kotin *et al.*, 1990). Thus, AAV has attracted great interest as a potential vector for stable gene delivery (Flotte & Carter, 1995; Monahan & Samulski, 2000). AAV packaging capacity is relatively small and, in most recombinant AAV (rAAV) vectors, only the 145 nt inverted terminal repeat (ITR) structures are retained. These can also facilitate genome integration, possibly through the action of cellular enzymes (Yang *et al.*, 1997), but in the absence of Rep, specificity for AAVS1 is lost and integration occurs at many chromosomal sites (Miller *et al.*, 2005; Nakai *et al.*, 2005).

Comprehensive knowledge of AAV behaviour is a prerequisite for the design of improved, targeted vectors.

While whole-genome strategies have been used to monitor integration by rAAV vectors (Miller *et al.*, 2005; Nakai *et al.*, 2005), for wild-type AAV (wtAAV), Southern blots and/or junction-specific PCR have traditionally been used, often using DNA from cell lines selected by neomycin resistance or long-term culture (Hamilton *et al.*, 2004; Huser *et al.*, 2002; Kotin *et al.*, 1990). In one study, fluorescent *in situ* hybridization was used to show that wtAAV integration occurred predominantly on chromosome 19, but also on chromosomes 1, 2 and 16 (Kearns *et al.*, 1996). In other studies, PCR methods were used to analyse integration events in human tissue, but these were confounded by abundant episomal material (Schnepp *et al.*, 2005) or by limited sample availability (Mehrlé *et al.*, 2004). Of three chromosome junctions identified, one and two were found on chromosomes 1 and 19, respectively.

New strategies, such as ligation-mediated PCR (LMPCR), were recently devised to analyse integration sites for retro- and lentiviruses at the whole-genome level. These sites are now known to be widely dispersed throughout the genome (Narezkina *et al.*, 2004; Schroder *et al.*, 2002; Wu *et al.*, 2003). The use of LMPCR to monitor whole-genome AAV

Supplementary material is available with the online version of this paper.

integration would therefore enable direct comparisons between wtAAV and retro-/lentiviruses and their current and future vector derivatives. To permit that, here we first studied AAV sequence integration in neomycin-selected cells by LMPCR, using a simple plasmid system to establish the methodology. We then analysed integration by wtAAV after infection in two unselected human cell lines, by coupling LMPCR with *SmaI* digestion to remove unintegrated episomal material. DNA was harvested from cells only a few days (4–11 days) after infection to simulate the short-term action of AAV in human or animal tissues.

For the plasmid system, we found, as expected, that viral p5-ITR elements may confer substantial Rep-mediated AAVS1 integration specificity on a selectable gene (Philpott *et al.*, 2002a, b). However, for wtAAV in unselected cells, we found that the virus (or parts thereof) can integrate specifically into chromosome 19 AAVS1 or join to other chromosomes via a few preferred 'GGTC' sequences that lie mainly within the 138 nt AAV p5 promoter. The overall complexity of wtAAV integration in unselected cells thus appears far greater than when only certain parts of that virus (e.g. Rep, p5 and ITR) are studied in a typical selectable cell system.

METHODS

Cell culture, AAV preparation and infection and plasmid transfection. Human cervical carcinoma (HeLa) and HepG2 liver carcinoma cells were obtained and cultured as previously described (Khatri *et al.*, 1997; Lockett & Both, 2002). Purified wtAAV2 (8×10^{10} infectious particles ml^{-1}) (Halbert *et al.*, 1997) was the generous gift of Dr Ian Alexander (Westmead Children's Hospital, Sydney, NSW, Australia). HeLa or HepG2 cells (2×10^5 per well) were infected at an m.o.i. of 100, 500 or 1000 infectious units per cell in 0.5 ml serum-free medium for 1 h at 37 °C. The inoculum was removed, 4 ml medium plus 10 % fetal calf serum was added and cells were incubated at 37 °C in 5 % CO₂ until harvested at 4–11 days post-infection (p.i.).

A plasmid containing the Rep78 gene was kindly provided by Dr M. Urabe (Jichi Medical School, Japan). Rep78 coding sequences were subcloned into a Bluescribe-derived plasmid (Stratagene), under the control of the Rous sarcomavirus promoter and the bovine growth hormone (BGH) polyadenylation signal. The Rep-expressing plasmid (typically 0.1 µg) and a second plasmid (typically 1.0 µg) that contained a wild-type or modified p5 promoter and/or ITR, together with a simian virus (SV40)/neomycin cassette, were transfected together into HeLa cells (approx. 1×10^6 cells) using Lipofectamine (Invitrogen) or a Tris-based cationic lipid (Cameron *et al.*, 1999), and neomycin-resistant colonies were selected.

Preparation of total cellular DNA and PCR analysis. Total DNA was extracted from cells at 4 and 11 days post-AAV infection or from plasmid-transfected, neomycin-selected cells by digestion with proteinase K in buffer containing SDS, and purified using Nucleospin columns (Machery-Nagel) according to the manufacturer's instructions. DNA was digested to completion with *Sau3AI* and heated at 65 °C to kill the enzyme, before ligation under standard conditions with an adaptor compatible with GATC sticky ends (adaptor sequences are in Supplementary Table S1, available in JGV Online). *SmaI* digestion was used when appropriate to minimize PCR amplification of episomal AAV genomes. Our LMPCR strategy is illustrated

in Fig. 1. Alternatively, genomic DNA was cut with *AluI* and *XmaI* overnight, and the *Sau3AI/AluI* upper and complementary *AluI* lower adaptors were ligated overnight in the presence of additional *AluI* and *XmaI*. The DNA ligase was then heat-killed, and the material re-digested with *AluI* and *XmaI* to completely remove unwanted *in vitro* ligation products (especially from *XmaI*; blunt-ended ligation of the *AluI* adaptor does not re-create an *AluI* site). LMPCR was carried out using $2 \times$ *Taq* polymerase Master Mix (Promega) and nested primers LP1–LP3 (all primer sequences are in Supplementary Table S1, available in JGV Online), which were complementary to the linker, together with nested primers Neo P1–P3 from the *neoR* gene, or AAV primers AL1–AL3 or AR1–AR3 (Table 1). For *AluI*-based LMPCR, primers LP1 and *AluI* AL1 were used for first-round PCR, with subsequent rounds carried out with *AluI* LP12/AL2, then LP2/AL3. To analyse sequences located more internally in the AAV genome, primers AB1–AB3, which are located between nt 979 and 1043, were used. In most cases, only two rounds of LMPCR were carried out, each comprising 30 cycles (95 °C for 15 s, 55 °C for 30 s, 72 °C for 2 min). However, a third round of 20 cycles was often used to confirm the authenticity of PCR products from round 2. AAVS1-specific junction PCR was carried out using chromosome 19 primers 1200, 1600 or 2400A, B or C together with appropriate AL1–AL3, AR1–AR3, AS1–AS3, AB1–AB3 or neomycin primers. PCR products from round 2 or 3 nested PCR were cloned using the pGEM-T Easy vector (Promega) and colonies with different-sized *EcoRI* inserts of approximately 200–1500 bp were selected for sequencing (performed at SUPAMAC, Camperdown, NSW, Australia). Sequence homologies were identified by BLAST searches at the NCBI web site.

RESULTS

Analysis of plasmid integration events in selected cells by LMPCR and AAVS1 junction PCR

To optimize LMPCR, we first used plasmid transfection to achieve Rep-mediated integration of p5-containing AAV sequences (Philpott *et al.*, 2002a, b). HeLa cells were transfected with two plasmids, one of which expressed Rep78. The other carried a neomycin-resistance gene, wild-type ITR and native p5 promoter (Fig. 2), or a modified AAV sequence such as a short ITR (Fig. 2b) plus a wild-type p5 (plasmid p15), or a modified p5 promoter where four bases were mutated to create a 'consensus' Rep-binding (RBS) site (plasmid p19) (Fig. 2c). These and related p5-ITR plasmids (not shown) had been prepared for other purposes. Post-transfection, neomycin-resistant colonies were selected by standard procedures. In the presence of Rep, the p5 element conferred an approximate 10-fold increase in neomycin-resistant colonies, compared with controls where the Rep plasmid was omitted, or there was no RBS, or where Rep was added but the plasmids contained ITRs alone (data not shown). DNA was prepared from pooled and/or single colonies, digested to completion with *Sau3AI* and subjected to LMPCR. As a negative control, seven neomycin-resistant colonies, selected following p7WT transfection in the absence of Rep, were also analysed. In two colonies, the junction was found between the p5/ITR and chromosomes 3 or 20. For five others, DNA amplification was very poor and/or the plasmid–chromosome junction could not be identified, presumably because it occurred far from the Neo primers. Southern blot analysis

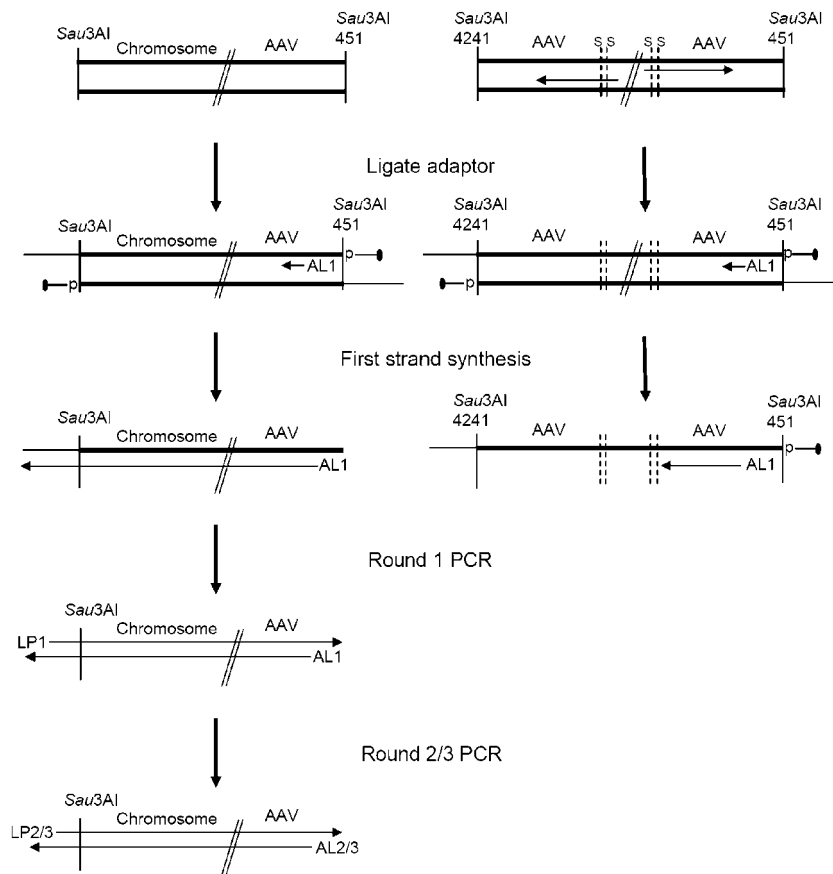


Fig. 1. Amplification of junction sequences by LMPCR. Our strategy follows Wu *et al.* (2003), except that *Sau3AI* was used instead of *Msel*. On the top line, the left side depicts an unknown AAV–chromosomal junction, while the right side shows a tail-to-head double-stranded concatemer of integrated or replicating AAV genomes, with known or potential *Sau3AI* and *SmaI* (S) sites and ITRs (long arrows). Alternative AAV configurations are also possible. On the second line, an *Sau3AI* adaptor is added by ligation. The black circle depicts a 3'-amino group that prevents extension of the adaptor during PCR. Thus, only when the first DNA strand primed by AL1 is complete can the LP1 priming site be created (line 3), so that first-round PCR (LP1/AL1) can be carried out (line 4). A cut at any *SmaI* site between AL1 and the nearest *Sau3AI* site then prevents unwanted amplification, as depicted on the right, for episomal AAV genomes. Random genomic *Sau3AI* fragments also will not be amplified, because the AL1 site is absent. Primers LP2/3 with AL2/3 are then used for round 2/3 PCR (line 5).

with an AAVS1 probe and selected genomic DNA confirmed that, in the absence of Rep, no chromosome 19 integration had occurred in any of these seven colonies (data not shown), in agreement with others (Philpott *et al.*, 2002b).

After transfection of p7WT, p15 and p19 plasmids in the presence of Rep, 305 LMPCR clones were grown and sequenced. Authentic cloned junctions (Recchia *et al.*, 2004) were defined as having: (i) appropriate, complete PCR primers present at one end or the other; (ii) chromosomal sequences that showed a unique match in a BLAST search against all organisms and a >95% match with the human chromosome; (iii) no *Sau3AI* sites at junctions between AAV and chromosomal sequences. We isolated 156 authentic junction clones with LP3 and Neo P3 primers. Of these, 34 (approx. 22%) contained AAV–chromosome 19 junctions. The break points lay within a 10–15 kb region that spanned, but lay mostly upstream of, the RBS in AAVS1 (Fig. 2d; Table 1; clones LM11-19 and LM226-15sc). Five clones were represented multiple times (a total of 15), the remainder were unique. In addition, 122 (78%) authentic junctions from other chromosomes were also cloned (examples are shown in Table 1). Most of these lay close to a putative, if 'weak', RBS in AAV or the chromosomal sequence (Table 1), thereby suggesting Rep-mediated events. Plasmid 7WT (wild-type ITR) and

modified plasmids p15/19 gave similar results (Table 1), indicating that the broad distribution of integration events (Fig. 2d) was not due to any ITR/p5 sequence modifications. When AAVS1-specific junction PCR was carried out on 7WT plasmid transfection samples using 1200, 1600 or 2400A, B or C and Neo P1–P3 primers, many junctions close to the chromosome 19 RBS were identified (Fig. 2d; Table 1), consistent with other observations (Philpott *et al.*, 2002b; Surosky *et al.*, 1997; Tsunoda *et al.*, 2000; Urabe *et al.*, 2003). Thus, whole-genome analysis of Rep-mediated plasmid integration events in neomycin-selected cells by LMPCR detected numerous authentic junctions (22%) in the general vicinity of AAVS1 on chromosome 19. Integration at sites on other chromosomes (78%) was also detected, most of which involved putative, if weaker, RBS. Integration of plasmid p5 sequences was also observed in Chinese hamster ovary cells (data not shown), consistent with the apparent ability of Rep to target non-chromosome 19 sites in mammalian cells in tissue culture.

Analysis of unselected AAV integration events p.i. using LMPCR and AAVS1-specific PCR

HeLa or HepG2 cells were infected with AAV, then DNA was harvested (without selection) at days 4 and 11 p.i. and completely digested with *Sau3AI*/*SmaI*. DNA samples were

Table 1. Location and sequences of LMPCR and AAVS1 junction clones derived from plasmid transfection and neomycin selection

RBS is at nt 27897132 in chromosome 19 (GenBank accession no. NT_011109.15, not in table). GenBank accession nos of sequences used are shown in italics.

Clone	Primers	Chromosome/GenBank accession no./position	Chromosomal junction sequence*†	AAV junction sequence*‡	AAV position
LM10-19	LP3 NeoP3	Chr 5, <i>NT_006576.15</i> , 30015806–30015867	<u>AGTGAGTGAGCGAGCGTG</u>	GGGCTTTGCCCGGGCGGCTC	69
LM11-19	LP3 NeoP3	Chr 19, <i>NT_011109.15</i> , 4965991–4965855	CCAATTGTGACTACTTAAT	GCGAGCGAGCGCGCAGAGAG	95
LM111-19	LP3 NeoP3	Chr 10, <i>NT_033985.6</i> , 2496199–2495820	AGGAGTCATCCAGGAGCAG	ACACCATGGGGTCACGCTGGG	234
LM108-15	LP3 NeoP3	Chr 16, <i>NW_926051.1</i> , 799916–799706	TCCAGCCTGGGTGAGTGAGCGAGCC	CGGCCTCAGTGAGCGAGCGAGCGCGC	83
LM194-15	LP3 NeoP3	Chr 9, <i>NT_023935.17</i> , 15727294–15727246	AACATTGAGCACTGAGTGAGCGAGC	GAGCGCGCAGAGAGGGAGTGGC	101
LM195-15	LP3 NeoP3	Chr 8, <i>NT_008046.15</i> , 44494743–44494640	CCAATCCTATTAGGAGCCTCCGTGTTT	GCCC GGCGGCCTCAGTGAGCGAGCGAG- CGCGC	76
LM226-15sc§	LP3 NeoP3	Chr 19, <i>NT_011109.15</i> , 27885410–27885331	AGCTAATTTTATTTTTCTAGAGACAA	CGGGCTTTGCCCGGGCGGCC	68
LM229-15sc	LP3 NeoP3	Chr X, <i>NT_011786.15</i> , 2757033–2756792	ACAAATAATCAGTCTCACTGTTACCAA	GCGAGCGCGCAGAGAGGGAGTGGCCA	99
LM43-7WT	LP3 NeoP3	Chr X, <i>NT_011651.16</i> , 2057201–2057166	TATTCTTTTGCTTCTGGATATTCAA	GGGGTGGAGTCGTGACGTGAATTAC	151
LM48-7WT	LP3 NeoP3	Chr 5, <i>NT_006576.15</i> , 30015867–30015824	GCAAGGCACAAAGCGGTGAGGGGAG	AGTGAGTGAGCGAGCGCGCAGAGAGG	90
LM144-7WT	LP3 NeoP3	Chr 5, <i>NT_006576.15</i> , 30026094–30026197	TGACAACACTTGCTTAACACATATAG	CCCGAGTGAGCACGCAGGGTCTCCATTT	263
J-7wt-501¶	1200B NeoP2	Chr 19, <i>NT_011109.15</i> , 27896487–27896713	GGATCGAGACTGGCAACGGGGAAGGAG	GTTAGGGAGGTCCTGTATTAGAGGTCAC	182
J-7wt-502	1200B NeoP2	Chr 19, <i>NT_011109.15</i> , 27896301–27896503	GTTAAAGCGACTCCAATGCGGAAG/ cgacctttggtcgcc/#	CGGCCTCAGTGAGCGAGCGAGCGCGCAG	83
J-7wt-507	2400B NeoP2	Chr 19, <i>NT_011109.15</i> , 27895128–27895551	GCCATGACAGGGGGCTGGAAGAGCTAG- CACAGA/gagc/	GAGCGAGCGCGCAGAGAGGGAGTGGCCA- ACTCC	93

*These sequences are joined tail to head to form the junction.

†Underlined nucleotides show homology to the AAVS1 RBS.

‡Bold letters indicate AAV RBS.

§Junction cloned from single-colony DNA.

||Three of 12 clones analysed are shown. Five were duplicates. Other independent junctions occurred on chromosomes 1, 5, 9 and 15.

¶J-clones were derived by AAVS1-specific junction PCR. Three of five clones analysed are shown. One was duplicated.

#Lower-case nucleotides are of undefined origin.

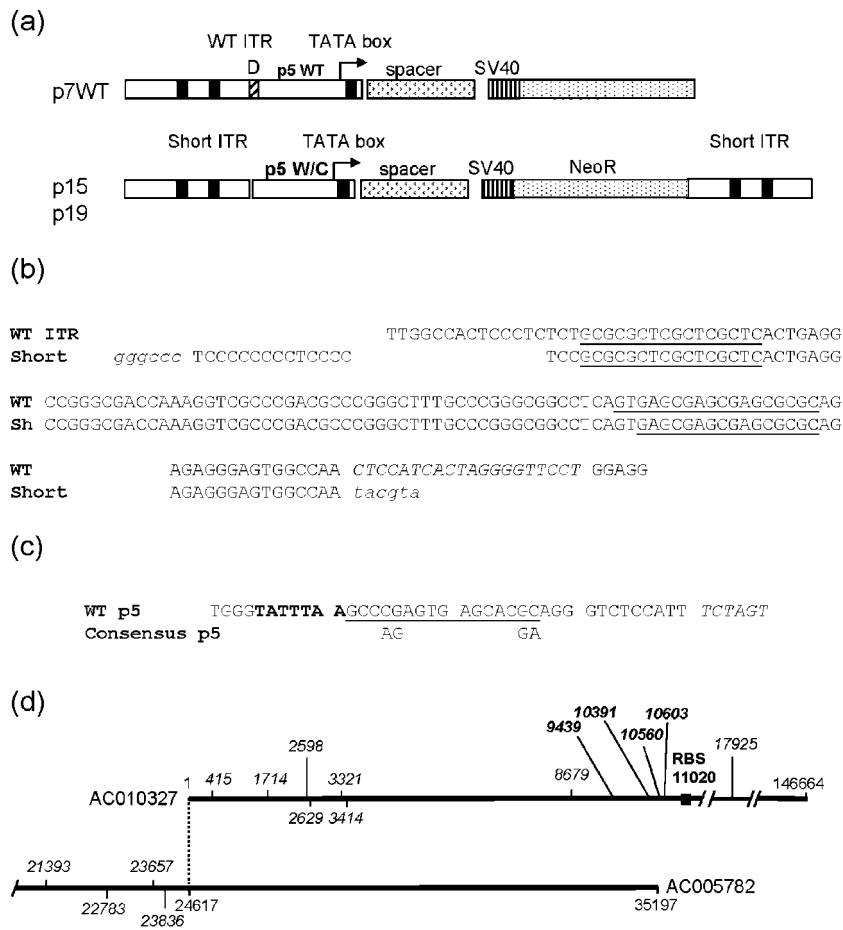


Fig. 2. Rep-mediated integration events in neomycin-selected cells after plasmid co-transfection. The structure of cassettes used in plasmids 7WT, p15 and p19 is shown in (a). Sequences are continuous, but gaps highlight differences between the ITRs. RBS are shown as black boxes in (a) and underlined in (b), where the D element is also shown in upper-case italics. Lower-case italics indicate restriction sites used in the construction. 7WT carries the wtAAV ITR sequence linked to an SV40 promoter/neomycin gene. Plasmids p15 and p19 carry a shortened ITR (b) linked to a wild-type (p15) or mutated (p19) p5 element (c). The TATA box is in bold type. In (d), the nucleotide number of cloned junctions in the chromosome 19 AAVS1 region is indicated on the maps of GenBank accession nos AC010327 (146 664 nt) and overlapping AC005782 (35 197 nt). Italicized numbers are LMPCR-derived clones. Bold italics indicate AAVS1-specific PCR junctions.

subjected to two or three rounds of LMPCR using primers AL1–AL3 or AR1–AR3 plus LP1–LP3 (Fig. 1), and analysed by gel electrophoresis (Fig. 3, lanes 2, 5, 9 and 12). Products were cloned and sequenced as described in Methods.

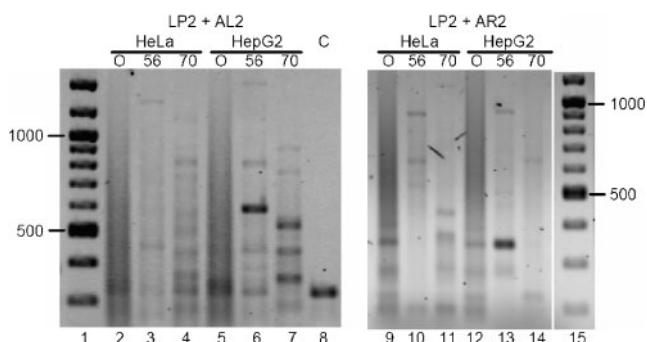


Fig. 3. Analysis of LMPCR products from unselected cells. DNA from HeLa or HepG2 cells infected with original (O) (lanes 2, 5, 9 and 12) or newly heated AAV (56 or 70 °C) was digested with *Sau3AI* and *SmaI*, and subjected to two rounds of LMPCR using primers LP2 and AL2 or AR2 in the second round. Products were analysed by agarose gel electrophoresis. Size markers (bp) are in lanes 1 and 15.

Very similar clones were isolated from HeLa cells at 4 or 11 days p.i. and also at an m.o.i. of 100 (17 clones day 4; 6 clones day 11) or 1000 (1 clone day 4; 10 clones day 11). Two clones came from HepG2 cells (m.o.i. of 100; day 4, where only a few clones were analysed). These data were therefore pooled (Fig. 4). Using the criteria described above, and excluding rare clones where a junction occurred at a *SmaI* site, we obtained 40 authentic junctions between the left end of AAV and human chromosomal sequences (Table 2). Four of those clones showed short AAV/chromosome sequence overlaps (micro-homologies) that were substantially different from each other, with AAV break points occurring at or near positions 195 (TGTATT), 337 (AGATT) and 344 (GATT) (numbering as per GenBank accession number J01901). A fourth clone had a sequence of 15 nt of unknown origin that linked position 282 to chromosome 16.

It was noteworthy that, in 36 of the 40 clones, sequences from various chromosomes were joined to AAV at or very near one of several GGTC motifs located at nucleotide positions 190, 204, 243 or 280 (Fig. 4a, b; Table 3). GGTC micro-homologies varied between 4 and 6 nt in length (two of 11 and 13 nt were also seen), and extended at most by 1 or 2 nt 5' to the motif. Position 280 was the most preferred motif (25 of 36 clones) (Fig. 4b; Table 3). In three clones

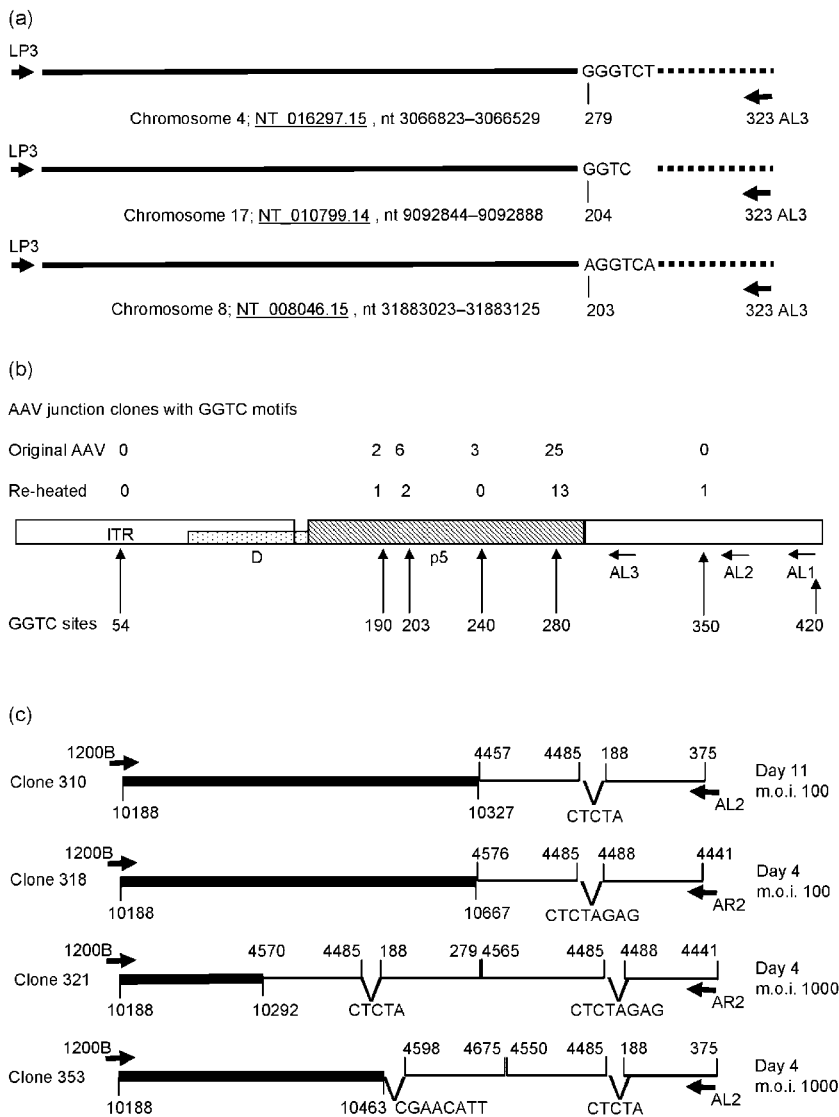


Fig. 4. Structure and location of AAV–chromosome junctions in unselected cells after AAV infection. (a) Structure of three typical GGTC clones derived from three rounds of LMPCR. The chromosome GenBank accession no. and nucleotides are shown along with the location of primers LP3 and AL3 and the nucleotide junction in or near a GGTC motif. (b) Structure of the left end of the AAV genome according to GenBank accession no. J01901 is depicted with its ITR, D and p5 elements, nested primers AL1–AL3 and the location of GGTC motifs. Above, the number of junctions found for each GGTC motif after infection with original or reheated AAV stock is shown. (c) Structure of four AAV–AAVS1 chromosome 19 junction clones. The ‘V’ represents the short non-AAV linker sequence shown in each case. Numbering as per GenBank accession no. AC010327.

Table 2. Summary of authentic cloned junctions derived from HeLa and HepG2 cells by LMPCR of unselected infected-cell DNA

Original AAV stock	Authentic junctions	
	GGTC micro-homology*	Other
Primers LP2/3 with		
AL2/3	36	4
AR2/3	0	12
Re-heated AAV stock, latent infection		
AL2/3	17	4
AR2/3	0	2

*Includes junction within or near the motif but with no sequence overlap.

there was no sequence overlap, but the break point occurred in or near the GGTC motif at positions 247, 279 and 282.

Although we sequenced fewer clones from the right end of AAV, junctions that were derived using primers AR2/AR3 did not show GGTC micro-homology (Table 2), but only two GGTC motifs exist in the last 282 nt of AAV. Right-end (five out of 12) clones showed various other micro-homologies that varied from 2 to 9 nt in length. Seven clones showed no sequence overlap. AAV sequences were linked to various chromosomes (e.g. 1, 5, 6, 8 and 10) via break points that occurred within the ITR between nt 4531 and 4593, as seen previously (Yang *et al.*, 1997; Miller *et al.*, 2002, 2004; Recchia *et al.*, 2004). One break occurred outside the ITR at position 4524.

None of these LMPCR clones obtained after AAV infection in the absence of selection contained an AAVS1 junction, whereas the neomycin-selected material above contained such junctions at a level of 22%. To confirm the existence

Table 3. Location of AAV motifs and break points in LMPCR cloned junctions from unselected infected-cell DNA

Position of GGTC motif*	No. of clones from unheated AAV	No. of clones from heated AAV
190	2 (1†, 11)	1 (5)
204	6 (3, 7, 8, 15, 17) 2 (2)	
243	3 (2, 8, 19)	0
280	25 (1, 2, 4, 5, 6, 8, 9, 11, 12, 13, 16, 21, X‡)	13 (1, 2, 10, 14, 16, 17, 20, 22)
350	0	1 (15)

*According to GenBank accession no. J01901.

†Chromosome number, in italics.

‡Multiple independent junctions were obtained for some chromosomes.

of AAVS1 junctions in unselected DNA, AL1/AL2 or AR1/AR2 and AAVS1 1200 or 1600A or B primers were used to amplify day 4 and day 11 genomic DNA. Of the clones sequenced, 23 out of 27 (85%) contained authentic chromosome 19 AAVS1 junctions (Fig. 4c). In contrast to Fig. 4(a), most of these clones showed rearranged or inverted AAV sequences typical of Rep-mediated, replication-dependent integration (Yang *et al.*, 1997), although there were a few clones in which fusion occurred without rearrangement. Since none of these were found in 52 examples of randomly picked clones from unselected DNA, their mean level is probably less than approximately 2%.

While our analysis of unselected clones was proceeding, we became aware that, despite having been heated during preparation, the AAV stock might still have a low level of contaminating AdVdl309 helper virus. This was confirmed when *EcoRV* digestion of DNA from infected cells produced a characteristic profile of AdV5 fragments on an agarose gel. Similarly, *BamHI* fragments of 3.6 and 1.05 kb indicated that the AAV genome had also replicated. These bands were already visible in HeLa cells at day 4 p.i., and were prominent at day 14 p.i. (Supplementary Figure S1, available in JGV Online). The AdVdl309 titre was determined directly in 293 cells at 8.3×10^6 TCID₅₀ ml⁻¹, implying an AdV m.o.i. of just 0.01 TCID₅₀ per cell, when cells were infected with an m.o.i. of 100. The conditions of AAV infection above may therefore be described as 'delayed-permissive', as there was too little helper virus for synchronous infection, but both genomes clearly replicated after a few days. The level of helper AdV required for synchronous infection (m.o.i. of 10–20 TCID₅₀ per cell) proved toxic to HeLa cells within 3 days (data not shown), precluding a meaningful analysis of integration under those conditions.

The original AAV stock was therefore reheated (at 70 °C for 10 min or 56 °C for 45 min) (Zhou & Muzyczka, 1998) to inactivate traces of AdVdl309, and tested using a sensitive

assay based on the development of a cytopathic effect (CPE) (Supplementary Figure S2, available in JGV Online). Cells infected with the original AAV stock developed signs of CPE by day 8 (m.o.i. of 5) or remained healthy at 12 days (low m.o.i. of 0.5), whereas all cells infected with newly heated stock (high m.o.i. of 500) remained healthy at 12 days p.i., clearly indicating that all AdVdl309 helper function had been inactivated. Newly heated AAV stock was therefore used to infect cells under true latent conditions.

Analysis of unselected AAV integration events after latent infection using LMPCR and AAVS1-specific PCR

HeLa and HepG2 cells were latently infected at an m.o.i. of 100 with newly heated AAV stock. Complete, unselected DNA was prepared, subjected to two or three rounds of LMPCR and analysed by gel electrophoresis. Compared with the samples from 'delayed-permissive' infection (Fig. 3, lanes 2, 5, 9 and 12), these second-round LMPCR products (Fig. 3, lanes 3 and 4, 6 and 7, 10 and 11, 13 and 14) were significantly less abundant and more discrete in size. Cloning and sequencing showed that some latent LMPCR products did not represent authentic junctions, yet authentic junctions were easily obtained from both cell types. Of 21 clones isolated, one recovered with right-end primers AR2/AR3 was joined to chromosome 8 with a TAATGAT micro-homology at position 4517 and three clones with AL2/LP3 primers were joined to chromosome 17 with AATCT micro-homology at position 337 (apparently the same clone). More significantly, 17 of the 21 clones isolated (81%) again showed GGTC micro-homology with break points at or near positions 280 (13), 203 (2), 190 (1) and 350 (1) (Fig. 4b; Table 3). These 17 clones were derived nearly equally from HepG2 and HeLa cells. Again, no chromosome 19 AAVS1 junctions were identified by LMPCR, but when junction-specific PCR was carried out with 1200 and 1600A or B primers, we isolated 12 authentic clones with junctions in the AAVS1 region. Thus, these data agree closely with those obtained in the first experiment, where a small amount of helper adenovirus was also present during infection.

To further quantify these data, numerous Southern blots were performed using LMPCR products that had been derived by plasmid co-transfection and neomycin selection (Fig. 2d), or unselected LMPCR products from infection with original or newly heated AAV stock (e.g. Fig. 3, lanes 2 and 3, respectively). As an internal control, authentic chromosome 19 junctions (such as in Fig. 4c) were also analysed. All samples were hybridized with an AAV- or a chromosome 19 AAVS1-specific probe. The relative strengths of the AAVS1/AAV signals obtained were found to be consistent with the identity of clones determined by direct sequencing (data not shown). It was estimated that, relative to an internal control, approximately 0.5–2% of LMPCR products from whole-genome, unselected DNA vs

10–20% of LMPCR products from plasmid-transfected, neomycin-selected DNA contained authentic chromosome 19 junctions (compared with 22% by direct sequencing). Other Southern blots were also performed using total genomic DNA isolated from single colonies of neomycin-selected cells (without PCR). In the presence of Rep, but not in its absence, 20–40% of those clones showed matching AAV and AAVS1 signals, consistent with most prior studies.

Thus, during both ‘delayed-permissive’ and latent infection by wtAAV and in the absence of any selection, at least two types of AAV–chromosomal junctions were observed: those involving GGTC micro-homology were apparently more numerous than those with chromosome 19 AAVS1 junctions.

Additional control experiments

We also conducted additional control experiments to determine whether any *in vitro* conditions for LMPCR might somehow favour GGTC over AAVS1 junctions among total LMPCR products.

A test for possible PCR sequence bias. We considered whether some unknown bias against amplification of sequences within AAVS1 (also a ‘CpG island’) could reduce the relative proportion of chromosome 19 junctions among total LMPCR clones. Separate mixtures of three GGTC and three AAVS1 junctions of distinct size were created from plasmid clones. Each clone contained at least one *Sau3AI* site in the chromosomal portion, either from the original linker ligation or because the DNA had not previously been cut with *Sau3AI* (AAVS1 clones isolated directly by PCR). Mixtures were spiked reciprocally in increasing dilutions into eight tubes containing a constant amount of human cellular DNA. Spiked samples were then digested with *Sau3AI* and ligated with the *Sau3AI* adaptor to precisely simulate LMPCR. One round of PCR was then performed using LP1 and AL2 primers. Under these conditions, all three AAVS1 junctions in each mixture amplified with similar efficiencies to the GGTC junctions (data not shown), indicating that primary sequence bias was not a major issue.

***AluI*- or *Bam*HI-based LMPCR.** We next conducted LMPCR using *AluI* or *Bam*HI to digest HeLa cell day 4–11 DNA, to determine whether our data might have somehow been influenced by an unexpected chromosomal site bias during *Sau3AI* digestion/ligation. Due to the lower efficiency of blunt-end ligation of the *AluI* adaptor, fewer authentic clones were isolated than with *Sau3AI*. But again, none of these contained junctions between AAV and chromosome 19 AAVS1. Clones with GGTC junctions at or near position 280 (two, linked to chromosomes 1 and 22) and 348 (one, linked to chromosome 17), similar to Fig. 4(a), were isolated with *AluI*. One clone with a break point at 259 (in the TATA box), linked to chromosome 1,

was also obtained. Three additional clones had a common structure where nt 323–188 of AAV were joined by a TAGAG linker to nt 4485–4593 of AAV, with the junction at 4593 joined to either chromosome 4, 13 or 19 (not in the AAVS1 region). The authentic *AluI* clones all came from latently infected cells.

Attempts to carry out LMPCR using *Bam*HI were confounded by an excess of viral episomal material that was not eliminated by digestion with *SmaI*, and probably also because the first *Bam*HI site on the left-hand side of AAV lies 1000 bp from *SmaI*, reducing PCR efficiency. No AAVS1-specific clones were obtained and only weak signals for AAVS1 were seen in Southern blots of LMPCR material. However, as a control, we easily obtained AAVS1-specific clones by junction PCR using a *Bam*HI site located approximately 1 kb upstream of the RBS in the AAVS1 region on chromosome 19. As *Bam*HI digestion at that site is not sensitive to DNA methylation and our LMPCR linker could have ligated at that junction, this strongly suggests that no AAV-S1 clones were obtained by LMPCR because they represented only a small fraction of the total material.

A test for junctions possibly formed by *in vitro* ligation.

We then investigated whether the low proportion of AAVS1 junctions among total LMPCR products could be caused by the possible ligation of compatible non-AAVS1 DNA fragments *in vitro*, rather than *in vivo*. Genomic DNA from AAV-infected cells was therefore treated with T4 DNA polymerase under standard conditions to remove protruding termini. The T4 enzyme was removed by heat treatment and silica column chromatography, blunt-ended DNA was digested with *Sau3AI/SmaI*, ligated with the *Sau3AI* adaptor and amplified by LMPCR. Analysis of T4-trimmed products and authentic control AAVS1 junctions by cloning/sequencing or by Southern blot showed that the percentage of AAVS1 junctions was still very low (0.5–2%), as observed above. We also isolated two authentic GGTC junction clones among the T4-trimmed clones that were sequenced.

Identification of AAV–chromosomal junctions by direct amplification from total unselected infected-cell DNA.

Lastly, if GGTC junctions are really produced *in vivo* after AAV infection, then by using primers based on chromosomal sequences determined from LMPCR, it should be possible to amplify specific junctions directly. We designed nested primers for four randomly selected GGTC junctions, amplified them directly by PCR from total, unselected genomic DNA, then cloned and sequenced the well-defined PCR products to confirm junction identity. In one case (clone 490, Table 4), we recovered the exact junction that was originally isolated (clone 941), albeit with different primers, and one variant (clone 942). In two other cases (original clones 211 and 212, Table 4), direct amplification produced authentic junctions in which the correct primers were present, but the break point in AAV and/or the chromosome varied due to the use of alternative GGTC

Table 4. Variations in AAV chromosomal junctions as seen by direct re-amplification of total unselected infected-cell DNA

GenBank accession nos of sequences used are shown in italics.

Original/new clone*	Primer	Chromosome/nucleotides	AAV nucleotides	AAV primer
211	LP3	Chromosome 8 <i>NT_008183.18</i>	247–375	AL2
		7480725–7480381†	No overlap‡	
935	Specific	7480688–7480375	280–323	AL3
			GGTC overlap	
830§	Specific	7480688–7480357	276–323	AL3
			GCAGGGTCT	
490	LP3	Chromosome 2 <i>NT_005403.16</i>	279–375	AL2
		51247044–51247364	GGGTCT	
941	Specific	51247076–51247364	279–323	AL3
			GGGTCT	
942	Specific	51247076–51247099	195–323	AL3
			No overlap	
212	LP3	Chromosome 10 <i>NT_008583.16</i>	195–375	AL2
		8426535–8426747	TGTATT	
937	Specific	8426531–8426548 then 14 nucleotide linker	188–323	AL3
			GAGGTC	

*Original clones are in bold.

†Last chromosomal nucleotide before any micro-homology.

‡Micro-homology between chromosome and AAV.

§Clone derived from *Sau3AI*-cut/ligated DNA.

motifs (clones 935, 942, 937, Table 4). One GGTC clone failed to amplify directly. Similar results were obtained by direct amplification of *Sau3AI*-cut/ligated genomic DNA with primers AL1–AL3 in combination with the respective chromosomal primers for four GGTC junctions (only clone 830 is shown). Thus, when measured directly, the true complexity of GGTC-mediated AAV–chromosomal junctions seems to be even greater than that determined from a subset of LMPCR clones.

DISCUSSION

A detailed knowledge of integration by wtAAV is required to design vectors with greater specificity. We therefore adapted strategies used by others (Recchia *et al.*, 2004; Wu *et al.*, 2003; Narezkina *et al.*, 2004; Schroder *et al.*, 2002) to monitor integration events in unselected infected-cell genomes. In an earlier study by Recchia *et al.* (2004), the absence of p5 sequences in the vector, plus the enzyme and primers chosen for LMPCR, precluded detection of the integration events described here. We chose *Sau3AI* rather than *MseI* (as used for other viruses) to cut genomic DNA, because there are no *Sau3AI* sites in the terminal 450 nt of the AAV genome at either end, thus allowing analysis of events across the whole ITR/p5 region (whereas *MseI* cuts in the AAV p5 TATA region). *Sau3AI* also cuts frequently within human chromosomal DNA near AAVS1, although half of those sites may be potentially blocked by overlapping CpG methylation.

To establish methods, we first analysed Rep-mediated integration of natural and modified AAV sequences, following co-transfection of two plasmids and selection for neomycin-resistant colonies using LMPCR and AAVS1 junction-specific PCR. Our data showing integration in the RBS of chromosome 19 were in agreement with others (Philpott *et al.*, 2002b; Surosky *et al.*, 1997; Tsunoda *et al.*, 2000; Urabe *et al.*, 2003) (Fig. 2d), but many junctions with other chromosomes were also identified (Table 1). Some chromosome 19 junctions flanked the RBS, but others were located 5–10 kb away, although in some bona fide chromatin structure they may not actually be widely spatially separated from the RBS. Most non-chromosome 19 junctions contained a putative RBS in their AAV or chromosomal sequence (bold and underlined letters, respectively; Table 1), suggestive of genuine Rep-mediated events rather than non-specific integration. In the absence of RSV/Rep, neomycin-resistant colony numbers were approximately 10-fold lower and only a few junctions in the neomycin plasmid could be identified, probably because they occurred outside of the AAV sequences and too far from the neomycin primers shown in Fig. 2.

We next examined integration across the whole genome during infection by wtAAV using LMPCR analysis of DNA from unselected infected-cell genomes. In the first experiment, ‘delayed-permissive’ conditions were effectively used because of trace levels of active helper virus in the original AAV stock. Analysis of DNA from cells at day 4 or 14 p.i. confirmed the replication of both the AAV and AdVdl309

genomes (Supplementary Figure S1, available in JGV Online). However, because the level of *Adv*l309 was initially very low (calculated as an m.o.i. of 0.01 when the AAV m.o.i. was 100), AAV integration in some cells may have initially occurred in the absence of AdV helper virus and essentially under latent conditions. This would have depended on how fast single- to double-stranded conversion and integration of AAV DNA occurred under latent conditions. For AAV in any context, this is a rate-limiting step (McCarty *et al.*, 2004). When all helper virus was removed by additional heat treatment, infection conditions were truly latent, as shown by a CPE assay. Our data show that, under both sets of conditions, the profiles of GGTC-mediated integration, as determined by LMPCR, were very similar (Fig. 4b).

In most analyses, we digested total genomic DNA with *Sau3AI* (or *AluI* or *BamHI*) and *SmaI*, the latter being essential to remove most of the concatenated and episomal sequences from the PCR. Because of this, viral-chromosomal junctions where a break point in AAV occurred external to a *SmaI* site in the terminal region of the genome are necessarily missing from our dataset of LMPCR clones (for example, we could not detect junctions at the GGTC motif, position 54). Yet, in most of our AAVS1 junction clones, the AAV break points were actually internal to those *SmaI* sites (for example, Fig. 4c). Thus, *SmaI* digestion should have a minimal impact on the ratio of AAVS1–non-AAVS1 junctions. There are also many *Sau3AI* and *AluI* sites located within the general AAVS1 region, plus a site for *BamHI* approximately 1 kb upstream. Thus, the isolation of few, if any, junctions would be compromised by digestion at a nearby *SmaI* site. Notably, even after *Sau3AI/SmaI* digestion, the break points that we identified by LMPCR near the right-hand end of AAV were similar to those described by others (Yang *et al.*, 1997). In fact, our data from the right end of AAV act as an internal control for data derived from the left end, where novel break points around GGTC sequences were observed. Thus, a broad distribution of integration sites across many chromosomes and a surprising prevalence of GGTC-mediated junctions over AAVS1 junctions were apparent after infection for all unselected infected cell samples.

Four additional control experiments were also carried out to determine whether any *in vitro* manipulations might have reduced the proportion of AAVS1 junctions among total LMPCR clones. We tested for possible bias against PCR amplification of the AAVS1 region and for inappropriate ligation of DNA fragments *in vitro*. We also showed that similar clones could be isolated using *AluI/XmaI* in an alternative LMPCR strategy (although the use of *BamHI* yielded only episomes). Finally, it was shown that cloned GGTC junctions, as identified by LMPCR, truly did exist in the whole genome, because they could be amplified directly by PCR using appropriate nested chromosomal primers. Indeed, the complexity of junctions revealed in this way was greater than for the limited number of clones identified by LMPCR (Table 4).

These new data reveal a greater complexity of AAV–chromosomal junctions than has been previously described, usually from studying material that has been neomycin-selected or expanded from a single cell over a long period. While it was not formally proven here that GGTC-mediated integration is Rep-dependent, these new observations correlate with the presence of the Rep gene, p5 element and an RBS in wtAAV, and were facilitated by the use of unselected, whole-genome LMPCR. Furthermore, the non-involvement of GGTC motifs at the right end of AAV and the lack of an RBS in that region also seem consistent. For neomycin-selected cells, the presence of sequences with homology to the RBS near other chromosomal junctions (Table 1) also suggests that Rep is involved. This new GGTC pathway appears to be mechanistically distinct from the non-specific integration events described for rAAV vectors (Miller *et al.*, 2004; Yang *et al.*, 1997). Those rAAV vectors lacked both Rep and p5 sequences, although micro-homologies were also involved in some integration events.

We did not see any GGTC junctions after plasmid transfection and long-term selection in the presence of Rep. Thus, it might be inferred that GGTC junction formation depends on the *in vivo* conversion of AAV DNA from its single- to double-stranded form. For example, the presence of Rep protein on its p5 RBS (bases 262–277) during the single- to double-strand filling process might explain why the GGTC motif at position 280 is most favoured, at least in the early stages of wtAAV infection. During the later stages of infection, once most conversion is complete, the well-known chromosome 19 AAV-S1 mechanism may dominate.

Although we detected left- and right-end integration events by LMPCR, we were unable to determine the long-range structure of integrated sequences, and therefore whether all or most of the 5 kb AAV genome might be present in any clone. We attempted to use a *BamHI* site at approximately position 1050 in AAV to generate longer left-end LMPCR products, but this was unsuccessful due to the prevalence of cloned episomal material.

Among our dataset of LMPCR clones, GGTC junctions commonly occurred in repeat DNA sequences; they occurred in only five cases within an intron and once within a coding region. Thus integration by wtAAV may differ from lenti- or retroviruses, which seem to preferentially target active chromatin or the genes within it (Narezkina *et al.*, 2004; Schroder *et al.*, 2002; Wu *et al.*, 2003).

In conclusion, whole-genome analysis of unselected infected-cell DNA has revealed new insights into integration by wtAAV. The detailed mechanism underlying the formation of GGTC-mediated junctions remains to be elucidated, but is outside the scope of this work. Finally, the specificity of Rep-dependent integration by new AAV vectors that may contain p5 sequences (Philpott *et al.*, 2002a, b; Recchia *et al.*, 2004) might usefully be monitored at the whole-genome level by methods such as those described here.

ACKNOWLEDGEMENTS

We thank A. Dalton and D. Lewy for assistance early in the project, M. Urabe for the Rep78 and other plasmids and I. Alexander for purified AAV2. The authors have no conflict of interest in relation to any of this work.

REFERENCES

- Cameron, F. H., Moghaddam, M. J., Bender, V. J., Whittaker, R. G., Mott, M. & Lockett, T. J. (1999). A transfection compound series based on a versatile Tris linkage. *Biochim Biophys Acta* **1417**, 37–50.
- Flotte, T. R. & Carter, B. J. (1995). Adeno-associated virus vectors for gene therapy. *Gene Ther* **2**, 357–362.
- Halbert, C. L., Standaert, T., Aitken, M., Alexander, I., Russell, D. & Miller, A. (1997). Transduction by adeno-associated virus vectors in the rabbit airway: efficiency, persistence, and readministration. *J Virol* **71**, 5932–5941.
- Hamilton, H., Gomos, J., Berns, K. I. & Falck-Pedersen, E. (2004). Adeno-associated virus site-specific integration and AAVS1 disruption. *J Virol* **78**, 7874–7882.
- Huser, D., Weger, S. & Heilbronn, R. (2002). Kinetics and frequency of adeno-associated virus site-specific integration into human chromosome 19 monitored by quantitative real-time PCR. *J Virol* **76**, 7554–7559.
- Kearns, W. G., Afione, S., Fiulmer, S., Pang, M., Erikson, D., Egan, M., Landrum, M., Flotte, T. & Cutting, G. (1996). Recombinant adeno-associated virus (AAV-CFTR) vectors do not integrate in a site-specific fashion in an immortalized epithelial cell line. *Gene Ther* **3**, 748–755.
- Khatri, A., Xu, Z. Z. & Both, G. W. (1997). Gene expression by atypical recombinant ovine adenovirus vectors during abortive infection of human and animal cells in vitro. *Virology* **239**, 226–237.
- Kotin, R. M., Siniscalco, M., Samulski, R., Zhu, X., Hunter, L., Laughlin, C., McLaughlin, S., Muzyczka, N., Rocchi, M. & Berns, K. (1990). Site-specific integration by adeno-associated virus. *Proc Natl Acad Sci U S A* **87**, 2211–2215.
- Lockett, L. J. & Both, G. W. (2002). Complementation of a defective human adenovirus by an otherwise incompatible ovine adenovirus recombinant carrying a functional E1A gene. *Virology* **294**, 333–341.
- McCarty, D. M., Young, S. M. & Samulski, R. J. (2004). Integration of adeno-associated virus (AAV) and recombinant AAV vectors. *Annu Rev Genet* **38**, 819–845.
- Mehrle, S., Rohde, V. & Schlehofer, J. R. (2004). Evidence of chromosomal integration of AAV DNA in human testis tissue. *Virus Genes* **28**, 61–69.
- Miller, D. G., Rutledge, E. A. & Russell, D. W. (2002). Chromosomal effects of adeno-associated virus vector integration. *Nat Genet* **30**, 147–148.
- Miller, D. G., Petek, L. M. & Russell, D. W. (2004). Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat Genet* **36**, 767–773.
- Miller, D. G., Trobridge, G. D., Petek, L. M., Jacobs, M. A., Kaul, R. & Russell, D. W. (2005). Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. *J Virol* **79**, 11434–11442.
- Monahan, P. E. & Samulski, R. J. (2000). AAV vectors: is clinical success on the horizon?. *Gene Ther* **7**, 24–30.
- Nakai, H., Wu, X., Fuess, S., Storm, T. A., Munroe, D., Montini, E., Burgess, S. M., Grompe, M. & Kay, M. A. (2005). Large-scale molecular characterization of adeno-associated virus vector integration in mouse liver. *J Virol* **79**, 3606–3614.
- Narezkina, A., Taganov, K. D., Litwin, S., Stoyanova, R., Hayashi, J., Seeger, C., Skalka, A. M. & Katz, R. A. (2004). Genome-wide analyses of avian sarcoma virus integration sites. *J Virol* **78**, 11656–11663.
- Philpott, N. J., Giraud-Wali, C., Dupuis, C., Gomos, J., Hamilton, H., Berns, K. I. & Falck-Pedersen, E. (2002a). Efficient integration of recombinant adeno-associated virus DNA vectors requires a p5-rep sequence in cis. *J Virol* **76**, 5411–5421.
- Philpott, N. J., Gomos, J., Berns, K. I. & Falck-Pedersen, E. (2002b). A p5 integration efficiency element mediates Rep-dependent integration into AAVS1 at chromosome 19. *Proc Natl Acad Sci U S A* **99**, 12381–12385.
- Recchia, A., Perani, L., Sartori, D., Olgiati, C. & Mavilio, F. (2004). Site-specific integration of functional transgenes into the human genome by adeno/AAV hybrid vectors. *Mol Ther* **10**, 660–670.
- Schnepf, B. C., Jensen, R. L., Chen, C.-L., Johnson, P. R. & Clark, K. R. (2005). Characterization of adeno-associated virus genomes isolated from human tissues. *J Virol* **79**, 14793–14803.
- Schroder, A. R. W., Shinn, P., Chen, H., Berry, C., Ecker, J. R. & Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521–529.
- Surosky, R. T., Urabe, M., Godwin, S., McQuiston, S., Kurtzman, G., Ozawa, K. & Natsoulis, G. (1997). Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. *J Virol* **71**, 7951–7959.
- Tsunoda, H., Hayakawa, T., Sakuragawa, N. & Koyama, H. (2000). Site-specific integration of adeno-associated virus-based plasmid vectors in lipofected HeLa cells. *Virology* **268**, 391–401.
- Urabe, M., Kogure, K., Kume, A., Sato, Y., Tobita, K. & Ozawa, K. (2003). Positive and negative effects of adeno-associated virus Rep on AAVS1-targeted integration. *J Gen Virol* **84**, 2127–2132.
- Wu, X., Li, Y., Crise, B. & Burgess, S. M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**, 1749–1751.
- Yang, C. C., Xiao, X., Zhu, X., Ansardi, D., Epstein, N., Frey, M., Matera, A. & Samulski, R. (1997). Cellular recombination pathways and viral terminal repeat hairpin structures are sufficient for adeno-associated virus integration in vivo and in vitro. *J Virol* **71**, 9231–9247.
- Zhou, X. & Muzyczka, N. (1998). In vitro packaging of adeno-associated virus DNA. *J Virol* **72**, 3241–3247.